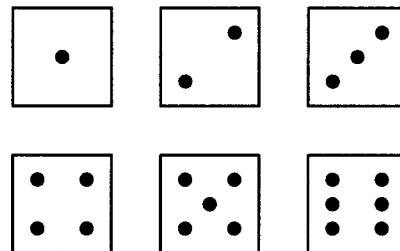


MODULE 1

PROBABILITY AND RANDOM VARIABLES

Probability Space

- A PROBABILITY SPACE is made up of a sample space, a set of events, and a probability measure.
- Central to the concept of probability space is an EXPERIMENT, the outcome of which cannot be known *a priori* with certainty.
- The SAMPLE SPACE \mathcal{S} is the set of all possible outcomes of the experiment.
 - ▶ Ex: If the experiment is the single throwing of a fair die, then \mathcal{S} consists of the six possible outcomes:



- ▶ Ex: If the experiment is to measure the temperature in Norman at 2:00 in the afternoon, then $\mathcal{S} = \mathbb{R}$, the set of real numbers.
- Note: the elements of \mathcal{S} must be disjoint. That is, for any trial of the experiment, the outcome can be one and only one element of \mathcal{S} .

- The SET OF EVENTS \mathcal{G} is a set whose members are all possible subsets of \mathcal{S} .
 - ▶ The empty set \emptyset and the set \mathcal{S} itself are both members of \mathcal{G} .
 - ▶ Members of \mathcal{G} can generally be constructed by taking unions, intersections, and complements of members of \mathcal{S} .
 - ▶ Formally, \mathcal{G} is known as a σ -algebra, but we will not concern ourselves with the details of this.

Ex: The event “the temperature in Norman at 2:00 is in the interval [72, 96) could be an event in \mathcal{G} (temperature experiment).

Ex: The event NOT((one dot) OR (two dots)) could be an event in \mathcal{G} (die experiment).

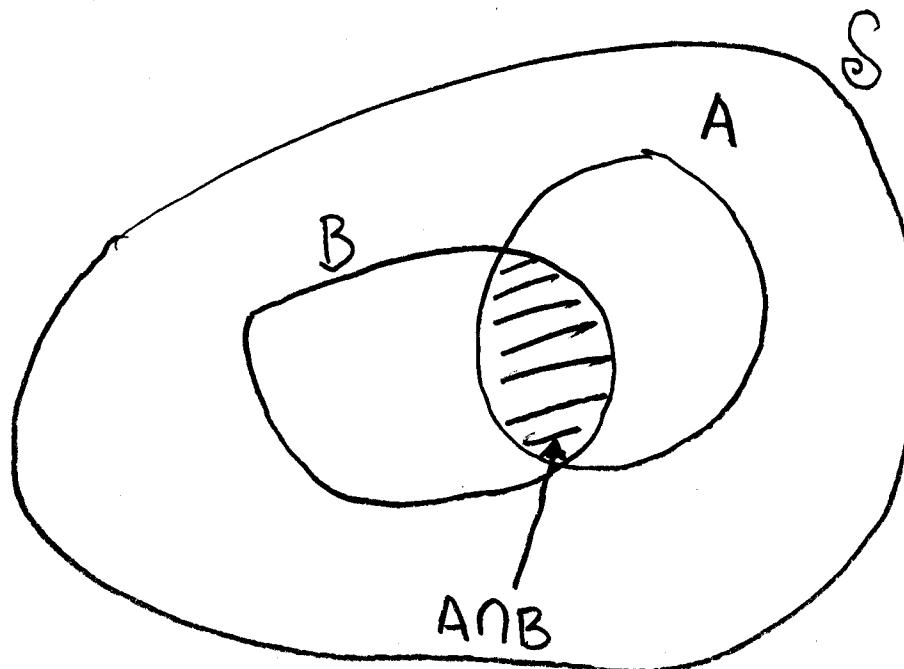
- ▶ Two events $A, B \in \mathcal{G}$ are mutually exclusive if $A \cap B = \emptyset$.

- The PROBABILITY MEASURE P is a function with domain is \mathfrak{S} and range $[0, 1]$.
 - Intuitively, we think of the number $P(A)$ as the probability that the event A occurs as the outcome of a random trial of the experiment.
- P must satisfy the following three axioms:
 1. $P(A) \geq 0 \forall A \in \mathfrak{S}$.
 2. $P(\mathcal{S}) = 1$.
 3. If $\{A_i\}$ is a countable collection of mutually exclusive sets in \mathfrak{S} , then
$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i).$$
- The following three properties are among the consequences of the three axioms:
 1. $P(A^C) = 1 - P(A)$.
 2. $P(\emptyset) = 0$.
 3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- Together, \mathcal{S} , \mathfrak{S} , and P makeup the PROBABILITY SPACE $(\mathcal{S}, \mathfrak{S}, P)$.

Notes

- For the temperature experiment, suppose that T is the measured temperature and that, on a given day, all temperatures between 75°F and 95°F are equally likely.
 - ▶ If A is the event $T = 80^{\circ}$ then $P(A) = 0$, since there are an infinite number of possible temperatures in the range $[75, 95]$.
 - ▶ For the event $A = \{T \in [75, 85)\}$, however, the probability is nonzero. In fact, $P(A) = 0.5$ in this case.
- For the die experiment, suppose that the die is fair, so that $P(\text{one dot}) = P(\text{two dots}) = \dots = P(\text{six dots}) = \frac{1}{6}$.
 - Let $A = \text{odd number of dots} = \boxed{\cdot} \cup \boxed{:} \cup \boxed{\bullet\bullet}$
 - $B = \text{number of dots} < 4 = \boxed{\cdot} \cup \boxed{:} \cup \boxed{:}$
 - $P(A) = P(B) = \frac{1}{2}$
 - $P(A \cap B) = P(\boxed{\cdot} \cup \boxed{:}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$
 - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 $= \frac{1}{2} + \frac{1}{2} - \frac{1}{3} = \frac{2}{3}$

CONDITIONAL PROBABILITY



$$\begin{aligned}
 - P(B|A) &= \text{"probability of } B \text{ given } A" \\
 &= \frac{P(B \cap A)}{P(A)}.
 \end{aligned}$$

BAYES' THEOREM :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

THEOREM OF TOTAL PROBABILITY:

Let $\{A_i\}$ be exhaustive ($\bigcup A_i = S$) and mutually exclusive. Then

$$P(B) = \sum_i P(B|A_i) P(A_i)$$

-COROLLARY:

$$P(A_k|B) = \frac{P(B|A_k) P(A_k)}{\sum_i P(B|A_i) P(A_i)}$$

DEF: Events A and B are called independent if
 $P(A \cap B) = P(A) P(B)$.

This implies that

$$P(A|B) = P(A)$$

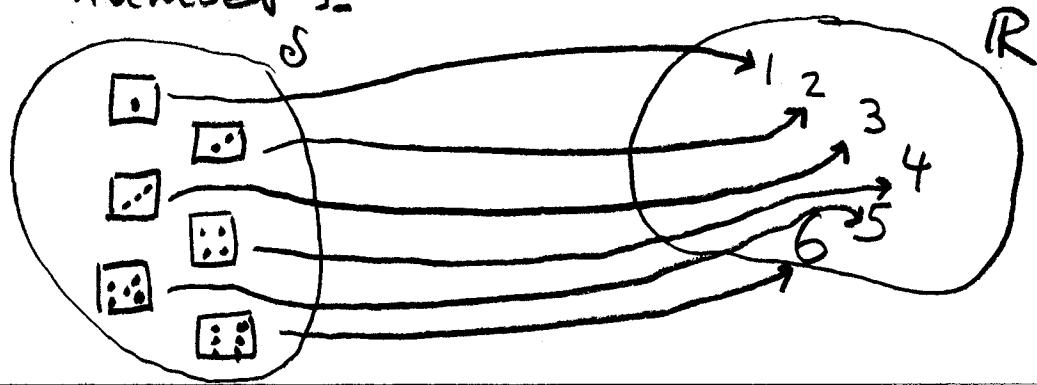
$$P(B|A) = P(B).$$

-Events that are not independent are called dependent.

RANDOM VARIABLES

- A random variable (RV) is a function with domain S and range \mathbb{R} .
- An RV "maps" the experimental outcomes to real numbers.

EX:



- Once the correspondence between experimental outcomes and real numbers has been made using an RV, the original sample space can often be ignored.
- This gives us a consistent way to treat a variety of experiments with a common mathematical framework.
- An RV for a sample space with a finite number of experimental outcomes is called a discrete RV.
- An RV for a sample space with an infinite number of experimental outcomes is called a continuous RV.

Cumulative Distribution Function (cdf) for Continuous RV

- For an RV X , the cdf is defined as

$$F_X(x) = P(X \leq x)$$

- The cdf is also referred to simply as the probability "distribution" function.

Properties:

1. $\lim_{x \rightarrow -\infty} F_X(x) = 0$
2. $\lim_{x \rightarrow \infty} F_X(x) = 1$
3. $F_X(x)$ is nondecreasing in x .
4. $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$.

Probability Density Function (pdf) for Continuous RV

- The pdf, or probability "density" is defined by

$$f_x(x) = \frac{d}{dx} F_x(x).$$

Properties:

1. $F_x(x) = \int_{-\infty}^x f_x(\theta) d\theta$

2. $P(x_1 < X \leq x_2) = F_x(x_2) - F_x(x_1) = \int_{x_1}^{x_2} f_x(\theta) d\theta$

3. $\int_{-\infty}^{\infty} f_x(\theta) d\theta = 1$

4. $f_x(x) \geq 0.$

Expected Value (mean) for a Continuous RV

- The mean of an RV X is written \bar{X} , $E[X]$, Ex , $\langle x \rangle$, m_x , or μ .

- By definition,

$$E[X] = \int_{-\infty}^{\infty} \theta f_x(\theta) d\theta.$$

Functions of a Continuous RV

- Suppose $g: \mathbb{R} \rightarrow \mathbb{R}$ is a function.

- Then we can consider $y = g(X)$, a function of the RV X .

- The RV X maps an experimental outcome $s \in S$ to a number $x \in \mathbb{R}$. The function g then maps x to a second number $y \in \mathbb{R}$.

- In symbols:

$$x \xrightarrow{X} x \xrightarrow{g} y$$

- The expected value, or mean, of $g(X)$ is given by

$$E[g(X)] = \int_{-\infty}^{\infty} g(\theta) f_X(\theta) d\theta.$$

Note: Since the expectation integral is a linear operator, we have that, with $\alpha, \beta \in \mathbb{R}$,

$$E[\alpha g_1(x) + \beta g_2(x)] = \alpha E[g_1(x)] + \beta E[g_2(x)].$$

Moments of a Continuous RV

- The " k^{th} moment" of an RV X is defined by

$$E[X^k] = \int_{-\infty}^{\infty} \theta^k f_X(\theta) d\theta.$$

- The zeroth moment is always equal to 1.

- The first moment is identical to the mean.

Central Moments of a Continuous RV

- The " k^{th} central moment of the RV X is the expected value of the function $g(x) = (x - \bar{x})^k$:

$$E[(x - \bar{x})^k] = \int_{-\infty}^{\infty} (\theta - \bar{x})^k f_X(\theta) d\theta$$

Notes:

- The zeroth central moment is always equal to 1.
- The first central moment is always equal to zero.
- The second central moment is called the "variance" σ_x^2 :

$$\sigma_x^2 = E[(x - \bar{x})^2] = \int_{-\infty}^{\infty} (\theta - \bar{x})^2 f_x(\theta) d\theta \\ = E[x^2] - (E[x])^2.$$

- The square root of the variance is called the "standard deviation" σ_x :

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{E[(x - \bar{x})^2]}$$

- The third central moment divided by the cube of the standard deviation is called the "skewness" of X :

$$\text{skewness} = \frac{E[(x - \bar{x})^3]}{\sigma_x^3}$$

- The fourth central moment divided by the square of the variance is called the "kurtosis" of X :

$$\text{kurtosis} = \frac{E[(x - \bar{x})^4]}{\sigma_x^4}$$

Characteristic Function of a Continuous RV

- The characteristic function of an RV X is given by

$$\psi_X(\omega) = \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx = E[e^{j\omega X}]$$

- Note:

$$\psi_X(-\omega) = \mathcal{F}\{f_X(x)\} ; f_X(x) = \mathcal{F}^{-1}\{\psi_X(-\omega)\}$$

- The pdf can be recovered from $\psi_X(x)$ by

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_X(\omega) e^{-j\omega x} d\omega$$

Moment Generating Function of a Continuous RV

- We can expand the characteristic function in a power series:

$$\psi_X(\omega) = E \left[\sum_{k=0}^{\infty} \frac{(j\omega x)^k}{k!} \right] = \sum_{k=0}^{\infty} \frac{(j\omega)^k}{k!} E[X^k]$$

- Furthermore,

$$\begin{aligned} \frac{d^k}{dw^k} \psi_X(w) \Big|_{w=0} &= \left[\frac{d^k}{dw^k} \int_{-\infty}^{\infty} f_X(x) e^{j\omega x} dx \right]_{w=0} \\ &= \left[\int_{-\infty}^{\infty} (jx)^k f_X(x) e^{j\omega x} dx \right]_{w=0} \\ &= j^k \int_{-\infty}^{\infty} x^k f_X(x) dx \\ &= j^k E[X^k] \end{aligned}$$

- The function $\phi_X(\omega) = \mathcal{F}_X\left(\frac{\omega}{j}\right)$ is called the "moment generating function" of the RV X .
- The moments of X can be obtained from the moment generating function by

$$E[X^k] = \left[\frac{d^k}{d\omega^k} \phi_X(\omega) \right]_{\omega=0}$$

Gaussian or "Normal" Random Variable

- The Gaussian distribution is used often because it approximately describes a large class of naturally occurring random phenomena.
- The terms "Gaussian" and "Normal" are used interchangeably.
- A Gaussian variable is described by two parameters :
 - The mean μ .
 - The variance σ^2 .
- The pdf of a Gaussian variable X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

- The cdf of a Gaussian variable is given by

$$\begin{aligned}
 F_X(x) &= \int_{-\infty}^x f_X(x) dx \\
 &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
 &= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\theta^2\right] d\theta \\
 &= \Phi\left(\frac{x-\mu}{\sigma}\right),
 \end{aligned}$$

where $\Phi(x)$ is the cdf of a Gaussian variable with zero mean and unit variance.

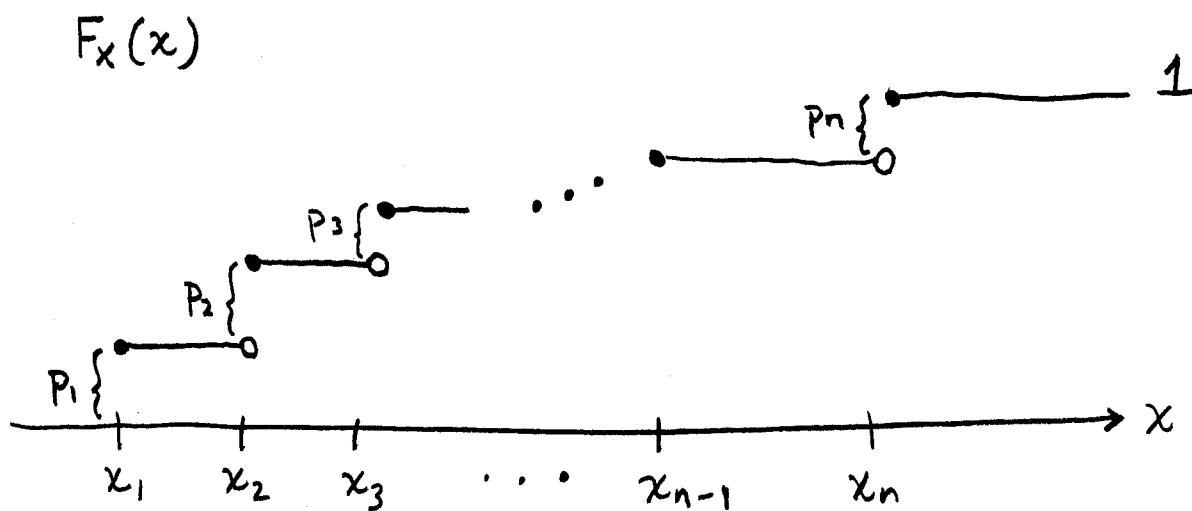
- The function $\Phi(x)$ cannot be expressed in terms of elementary functions. To evaluate it, we must use numerical integration or tables.
- Be Very Careful when using tables. Sometimes they are normalized in strange ways or specified in terms of the closely related "error function" $\text{E}_e(x)$.
- The shorthand notation $X \sim N(\mu, \sigma^2)$ is often used to indicate that the RV X is normally distributed with mean μ and variance σ^2 .
- The characteristic function for a $N(\mu, \sigma^2)$ variable is given by $\chi_X(w) = \exp\left[j\mu w - \frac{1}{2}\sigma^2 w^2\right]$

Discrete Random Variables



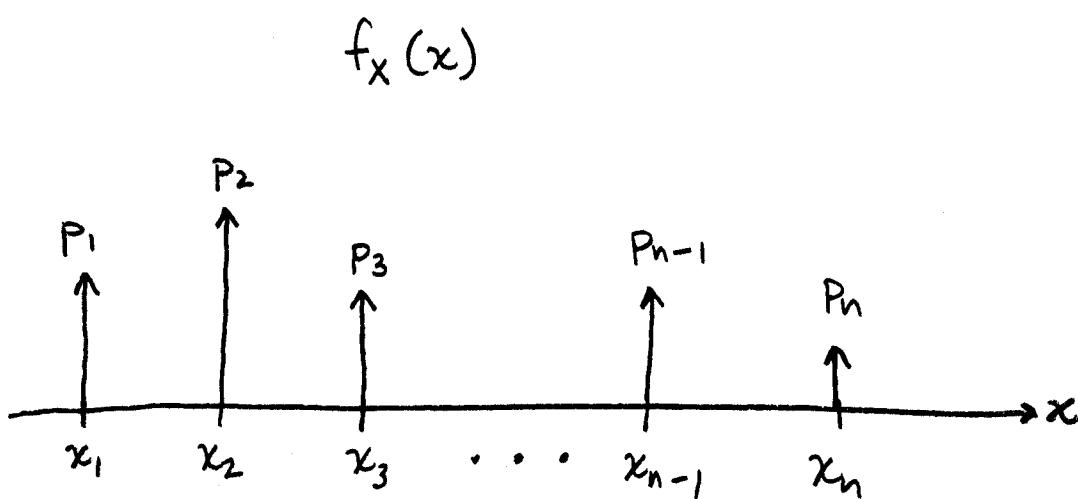
- Let $(\mathcal{S}, \mathcal{G}, P)$ be a probability space
- Suppose that the number of elements in \mathcal{S} is finite. The RV X maps the outcomes to a finite set of numbers $\{x_1, x_2, \dots, x_n\}$, where $x_1 < x_2 < \dots < x_n$.
- Then X is a discrete RV.
- Let $P(x_k) = p_k$.
- Then $\sum_{k=1}^n p_k = 1$.
- The cdf is given by

$$F_X(x) = \begin{cases} 0, & x < x_1 \\ \sum_{k=1}^i p_k, & x_i \leq x < x_{i+1} \\ 1, & x > x_n \end{cases}$$



- Taking the derivative of the cdf, we obtain a sum of weighted Dirac deltas for the pdf:

$$f_x(x) = \frac{d}{dx} F_x(x) = \sum_{k=1}^n p_k \delta(x - x_k)$$



- The mean is given by

$$E[X] = \int_{-\infty}^{\infty} x f_x(x) dx = \sum_{k=1}^n p_k x_k$$

- The variance is given by

$$\sigma^2 = \int_{-\infty}^{\infty} (x - E[X])^2 f_x(x) dx = \sum_{k=1}^n (x_k - E[X])^2 p_k$$

- The characteristic function is given by

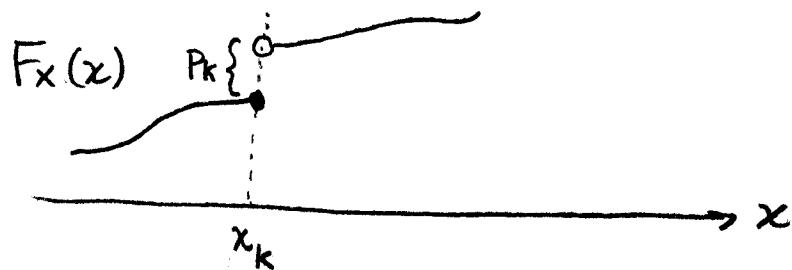
$$\psi_x(\omega) = \sum_{k=1}^n p_k e^{i\omega x_k}$$

- The expected value of the function $g(x)$ is given by

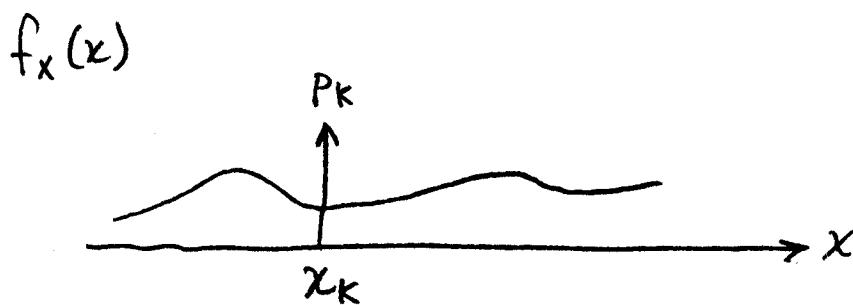
$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_x(x) dx = \sum_{k=1}^n g(x_k) p_k$$

Mixed Random Variables

- Suppose X is an RV that takes uncountably many values $x \in \mathbb{R}$, but that a finite set of these values $\{x_1, x_2, \dots, x_n\}$ each occur with nonzero probability.
- Then X is a "mixed" RV. The cdf and pdf have characteristics of both continuous and discrete RVs.
- Let $P(x_k) = p_k$.
- Note that $P(x) = 0$ for $x \neq x_k$, $k = 1, 2, \dots, n$.
- In this case, the cdf $F_x(x)$ has a step discontinuity of height p_k at x_k :



- The pdf has a Dirac delta of weight p_k at x_k :



Conditional Density

- The conditional distribution for the RV X given that event A has occurred is given by

$$F_{x|A}(x) = P(X \leq x | A)$$

- The corresponding conditional density is given by

$$f_{x|A}(x) = \frac{d}{dx} F_{x|A}(x)$$

- The corresponding event-on-density conditional probability is given by

$$P(A | X=x) = \lim_{\Delta x \rightarrow 0} P(A | x \leq X \leq x + \Delta x)$$

- The Bayes' formula relating the conditional density and conditional probability is

$$f_{x|A}(x) = \frac{P(A | X=x) f_x(x)}{P(A)}$$

Multiple Random Variables

- "Bivariate" refers to a situation involving two RVs,
- "Multivariate" refers to a situation involving two or more RVs,

Joint Distribution:

- The joint cdf of the n RVs x_1, x_2, \dots, x_n is given by

$$F_{x_1, \dots, x_n}(x_1, \dots, x_n) = P(x_1 \leq x_1 \cap x_2 \leq x_2 \cap \dots \cap x_n \leq x_n)$$

- Properties:

$$\lim_{x_1, x_2, \dots, x_n \rightarrow -\infty} F_{x_1, \dots, x_n}(x_1, \dots, x_n) = 0$$

$$\lim_{x_1, \dots, x_n \rightarrow \infty} F_{x_1, \dots, x_n}(x_1, \dots, x_n) = 1$$

Joint Density:

- The joint pdf of the n RVs x_1, x_2, \dots, x_n is given by

$$f_{x_1, \dots, x_n}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F_{x_1, \dots, x_n}(x_1, \dots, x_n)$$

- The joint cdf can be recovered from the joint pdf by integration:

$$F_{x_1, \dots, x_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_{x_1, \dots, x_n}(\theta_1, \dots, \theta_n) d\theta_1 \cdots d\theta_n$$

- The joint pdf integrates to 1:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{x_1, \dots, x_n}(\theta_1, \dots, \theta_n) d\theta_1 \cdots d\theta_n = 1$$

- The marginal density of the variable X_k may be obtained from the joint pdf by "integrating out" the other variables;

$$f_{x_k}(x_k) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-1 \text{ integrals}} f_{x_1, \dots, x_n}(x_1, \dots, x_n) dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n$$

$\underbrace{dx_k}_{\text{"dx}_k\text{" missing}}$

- The marginal cdf for X_k is then obtained by

$$F_{x_k}(x_k) = \int_{-\infty}^{x_k} f_{x_k}(\theta) d\theta$$

- The variables X_1, \dots, X_n are called mutually independent if

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{k=1}^n F_{X_k}(x_k)$$

- This is equivalent to

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{k=1}^n f_{X_k}(x_k)$$

→ i.e., they are independent if the joint cdf is the product of the marginal cdfs and the joint pdf is the product of the marginal pdfs.

Note: For discrete RVS, all of the above multivariate formulas involving integrals reduce to sums.

Functions of Multiple RVS/

- Suppose $g: \mathbb{R}^n \rightarrow \mathbb{R}$ is a function.
- Let X_1, \dots, X_n be n RVS/ with joint cdf $F_{X_1, \dots, X_n}(x_1, \dots, x_n)$ and joint pdf $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$.
- Then $g(X_1, \dots, X_n)$ is a function of the n RVS that takes scalar values in \mathbb{R} .

- The expected value of $g(x_1, \dots, x_n)$ is

$$E[g(x_1, \dots, x_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f_{x_1, \dots, x_n}(x_1, \dots, x_n) dx_1 \cdots dx_n$$

- If the RV's are mutually independent and g is separable so that $g(x_1, \dots, x_n) = \prod_{k=1}^n g_k(x_k)$, then

$$E[g(x_1, \dots, x_n)] = \prod_{k=1}^n E[g_k(x_k)]$$

EX: Suppose $g(x_1, \dots, x_n) = \sum_{k=1}^n \alpha_k x_k$ and x_1, \dots, x_n are mutually independent. Then

$$\begin{aligned} E[g(x_1, \dots, x_n)] &= E\left[\sum_{k=1}^n \alpha_k x_k\right] \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{k=1}^n \alpha_k x_k f_{x_1, \dots, x_n}(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \sum_{k=1}^n \alpha_k \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_k f_{x_1, \dots, x_n}(x_1, \dots, x_n) dx_1 \cdots dx_n \\ &= \sum_{k=1}^n \alpha_k \int_{-\infty}^{\infty} x_k f_{x_k}(x_k) dx_k \underbrace{\prod_{l=1, l \neq k}^n \int_{-\infty}^{\infty} f_{x_l}(x_l) dx_l}_1 \\ &= \sum_{k=1}^n \alpha_k \int_{-\infty}^{\infty} x_k f_{x_k}(x_k) dx_k \\ &= \sum_{k=1}^n \alpha_k E[X_k]. \end{aligned}$$

- We can also consider a vector-valued function $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$.

- In this case, $g(x_1, \dots, x_n)$ is the vector

$$g(x_1, \dots, x_n) = [y_1 \dots y_m]^T = \begin{bmatrix} g_1(x_1, \dots, x_n) \\ g_2(x_1, \dots, x_n) \\ \vdots \\ g_m(x_1, \dots, x_n) \end{bmatrix}$$

- To find the expected value of $g(x_1, \dots, x_n)$, we take the expectation on an element-by-element basis, and $E[g(x_1, \dots, x_n)]$ is the vector

$$E[g(x_1, \dots, x_n)] = E \begin{bmatrix} g_1(x_1, \dots, x_n) \\ g_2(x_1, \dots, x_n) \\ \vdots \\ g_m(x_1, \dots, x_n) \end{bmatrix}$$

$$= \begin{bmatrix} E[g_1(x_1, \dots, x_n)] \\ E[g_2(x_1, \dots, x_n)] \\ \vdots \\ E[g_m(x_1, \dots, x_n)] \end{bmatrix}$$

Joint Moments

- The joint moment of order k_1, \dots, k_n of the n RVs x_1, \dots, x_n is given by

$$\begin{aligned} E[x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}] &= E\left[\prod_{l=1}^n x_l^{k_l}\right] \\ &= \int_{\mathbb{R}^n} \prod_{l=1}^n x_l^{k_l} f_{x_1, \dots, x_n}(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

Joint Central Moments

- The joint central moment of order k_1, \dots, k_n of the n RVs x_1, \dots, x_n is given by

$$\begin{aligned} E\left[\prod_{l=1}^n (x_l - \bar{x}_l)^{k_l}\right] \\ = \int_{\mathbb{R}^n} \prod_{l=1}^n (x_l - \bar{x}_l)^{k_l} f_{x_1, \dots, x_n}(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

Covariance and Correlation for Two RVs

- Let X_1 and X_2 be two RVs with joint pdf $f_{X_1, X_2}(x_1, x_2)$.
- The covariance of X_1 and X_2 is the joint central moment
$$\text{Cov}(X_1, X_2) = \iint_{\mathbb{R}^2} (x_1 - \bar{x}_1)(x_2 - \bar{x}_2) f_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$
- Using the Cauchy-Schwartz inequality, it is easy to show that
$$|\text{Cov}(X_1, X_2)| \leq \sigma_{X_1} \sigma_{X_2}$$
- The correlation coefficient between X_1 and X_2 is given by
$$\rho_{X_1, X_2} = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}}$$
- Note that $|\rho_{X_1, X_2}| \leq 1$.
 - if $\rho=1$, then $X_1 = \alpha X_2$ with $\alpha > 0$. (Perfectly correlated)
 - if $\rho=-1$, then $X_1 = \alpha X_2$ with $\alpha < 0$. In this case, X_1 and X_2 are said to be anticorrelated.
 - if $\rho=0$, then X_1 and X_2 are said to be uncorrelated.

Notes:

- In general, $\text{Cov}(X_1, X_2) = E[X_1 X_2] - E[X_1] E[X_2]$

- If X_1 and X_2 are independent, then

$$E[X_1 X_2] = \int_{\mathbb{R}^2} x_1 x_2 f_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$

$$= \int_{\mathbb{R}^2} x_1 x_2 f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2$$

$$= \int_{\mathbb{R}} x_1 f_{X_1}(x_1) dx_1 \int_{\mathbb{R}} x_2 f_{X_2}(x_2) dx_2$$

$$= \bar{X}_1 \bar{X}_2$$

\Rightarrow So independence implies $\text{Cov}(X_1, X_2) = 0$

\Rightarrow This implies $\rho_{X_1, X_2} = 0$

\Rightarrow So independence implies uncorrelated.

$\rightarrow \star$ The converse is not true.

DEF: if $E[X_1 X_2] = 0$, then X_1 and X_2 are called orthogonal.

- If X_1 and X_2 are orthogonal and if $\bar{X}_1 = 0$ or $\bar{X}_2 = 0$, then X_1 and X_2 are uncorrelated.

Covariance for Random Vectors

- Suppose \vec{X} and \vec{Y} are random vectors (vectors of RVs) with

$$\vec{X} = [x_1 \dots x_n]^T \text{ and } \vec{Y} = [y_1 \dots y_m]^T.$$

- The covariance of \vec{X} and \vec{Y} is then a covariance matrix defined by

$$\begin{aligned}\text{Cov}(\vec{X}, \vec{Y}) &= \begin{bmatrix} \text{Cov}(x_1, y_1) & \text{Cov}(x_1, y_2) & \dots & \text{Cov}(x_1, y_m) \\ \text{Cov}(x_2, y_1) & \text{Cov}(x_2, y_2) & \dots & \text{Cov}(x_2, y_m) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(x_n, y_1) & \text{Cov}(x_n, y_2) & \dots & \text{Cov}(x_n, y_m) \end{bmatrix} \\ &= E[(\vec{X} - E[\vec{X}])(\vec{Y} - E[\vec{Y}])^T].\end{aligned}$$

- $\text{Cov}(\vec{X}, \vec{X})$ is called the covariance matrix of the random vector \vec{X} , denoted $\text{Cov}(\vec{X})$,

Note :

$$\text{Cov}(\vec{X}) = \begin{bmatrix} \sigma_{x_1}^2 & \rho_{x_1, x_2} \sigma_{x_1} \sigma_{x_2} & \dots & \rho_{x_1, x_n} \sigma_{x_1} \sigma_{x_n} \\ \rho_{x_2, x_1} \sigma_{x_2} \sigma_{x_1} & \sigma_{x_2}^2 & \dots & \rho_{x_2, x_n} \sigma_{x_2} \sigma_{x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{x_n, x_1} \sigma_{x_n} \sigma_{x_1} & \rho_{x_n, x_2} \sigma_{x_n} \sigma_{x_2} & \dots & \sigma_{x_n}^2 \end{bmatrix}$$

- The covariance matrix of a random vector is always positive semidefinite : for any vector \vec{a} ,

$$\vec{a}^T \text{Cov}(\vec{X}) \vec{a} \geq 0.$$

→ The eigenvalues of $\text{Cov}(\vec{X})$ are all nonnegative.

Note: The joint characteristic function and moment generating function for a random vector are easily generated using the multidimensional Fourier transform.

Conditional Density for Two RV's

- Let X and Y be two RV's with joint density $f_{X,Y}(x,y)$ and marginal densities $f_X(x)$ and $f_Y(y)$.

- Then

$$\begin{aligned} f_{X|Y=y}(x) &= \lim_{\Delta y \rightarrow 0} f_{X|\{y < Y \leq y + \Delta y\}}(x) \\ &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \end{aligned}$$

- This has meaning only if $f_Y(y) \neq 0$.

Sum of Two Independent RV's

- Let X and Y be two independent RV's with marginal densities $f_X(x)$ and $f_Y(y)$.

- Let Z be a third RV defined by $Z = X + Y$.

$$\begin{aligned} \text{- Then } f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-\theta) f_Y(\theta) d\theta \\ &= \int_{-\infty}^{\infty} f_X(\theta) f_Y(z-\theta) d\theta \\ &= f_X(z) * f_Y(z), \quad (\text{Convolution}) \end{aligned}$$

- In the book, this result is obtained using a standard argument.
- Here, we will take a different approach based on the characteristic function.
- Since X and Y are independent, we have that

$$\begin{aligned}
 \psi_z(\omega) &= E[e^{j\omega Z}] \\
 &= E[e^{j\omega(X+Y)}] \\
 &= E[e^{j\omega X} e^{j\omega Y}] \\
 &= E[e^{j\omega X}] E[e^{j\omega Y}] \quad (\text{because independent}) \\
 &= \psi_x(\omega) \psi_y(\omega)
 \end{aligned}$$

- So $\psi_z(-\omega) = \psi_x(-\omega) \psi_y(-\omega)$
- Taking the inverse Fourier Transform:

$$\begin{aligned}
 f_z(z) &= \mathcal{F}^{-1}\{\psi_z(-\omega)\} \\
 &= \mathcal{F}^{-1}\{\psi_x(-\omega) \psi_y(-\omega)\} \\
 &= f_x(z) * f_y(z)
 \end{aligned}$$

Transformations of an RV

- Let X be an RV and $g: \mathbb{R} \rightarrow \mathbb{R}$ be a function.

- Then we can define a new RV Y by the transformation

$$Y = g(X)$$

- If g is a "one-to-one" function (a bijection), then it is invertible and g^{-1} is also a function.

Ex:

$$y = g(x) = 5x + 3$$

$$x = g^{-1}(y) = \frac{y-3}{5}$$

- If g is one-to-one and $Y = g(X)$, then the density of Y is given by

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y))$$

\Rightarrow A derivation and examples are given in the book.

\Rightarrow If g is not one-to-one, it is generally necessary to resort to fundamental probability concepts to find $f_Y(y)$.

\Rightarrow An example multivariate transformation is also given in the book.

Jointly Gaussian Variables

- Let X_1, X_2, \dots, X_n be n RV's.

- Let \vec{X} be the random vector

$$\vec{X} = [x_1 \ x_2 \ \dots \ x_n]^T$$

- Let $\vec{m}_{\vec{X}}$ be the expected value of \vec{X} :

$$\begin{aligned}\vec{m}_{\vec{X}} &= E[\vec{X}] = [E[x_1] \ E[x_2] \ \dots \ E[x_n]]^T \\ &= [m_1 \ m_2 \ \dots \ m_n]^T\end{aligned}$$

- Let $C_{\vec{X}}$ be the covariance matrix of \vec{X} :

$$C_{\vec{X}} = \text{cov}(\vec{X}, \vec{X}) = E[(\vec{X} - \vec{m})(\vec{X} - \vec{m})^T]$$

- Let $|C_{\vec{X}}| = \det C_{\vec{X}}$

- The multivariate normal density is given by

$$f_{\vec{X}}(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |C_{\vec{X}}|^{1/2}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{m}_{\vec{X}})^T C_{\vec{X}}^{-1} (\vec{x} - \vec{m}_{\vec{X}}) \right\}$$

Note: $f_{\vec{X}}(\vec{x}) = f_{\vec{X}}(x_1, x_2, \dots, x_n) : \mathbb{R}^n \rightarrow [0, \infty)$

- In the bivariate case,

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \vec{m}_{\vec{X}} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$

$$C_{\vec{X}} = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}$$

where $\rho = \text{corr}(X_1, X_2)$

$$|C_{\vec{X}}| = \sigma_1 \sigma_2 (1 - \rho^2)$$

$$C_{\vec{X}}^{-1} = \begin{bmatrix} \frac{1}{(1-\rho^2)\sigma_1^2} & \frac{-\rho}{(1-\rho^2)\sigma_1 \sigma_2} \\ \frac{-\rho}{(1-\rho^2)\sigma_1 \sigma_2} & \frac{1}{(1-\rho^2)\sigma_2^2} \end{bmatrix}$$

- So

$$f_{\vec{X}}(\vec{x}) = f_{X_1, X_2}(x_1, x_2)$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(x_1 - m_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - m_1)(x_2 - m_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} \right] \right\}$$

- Suppose X_1 and X_2 are uncorrelated.

Note: this is a weaker assumption than independent.

- Then $\rho=0$ and $C_x = \text{diag}(\sigma_1^2 \ \sigma_2^2)$.

- So

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp \left\{ -\frac{1}{2} \left[\frac{(x_1-m_1)^2}{\sigma_1^2} + \frac{(x_2-m_2)^2}{\sigma_2^2} \right] \right\}$$
$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(x_1-m_1)^2/2\sigma_1^2} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(x_2-m_2)^2/2\sigma_2^2},$$

the product of the marginal densities.
(for X_1, X_2 uncorrelated).

⇒ Thus, for Gaussian variables,

Uncorrelated \longleftrightarrow independent

Linear Transformation of Jointly Gaussian Variables

- Let $\vec{X} = [X_1 X_2 \dots X_n]^T$ be a vector of n jointly Gaussian RVs,

- The multivariate density of \vec{X} is

$$f_{\vec{X}}(\vec{x}) = (2\pi)^{-\frac{n}{2}} |C_{\vec{X}}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} (\vec{x} - \vec{m}_{\vec{X}})^T C_{\vec{X}}^{-1} (\vec{x} - \vec{m}_{\vec{X}})\right\}$$

- Define a random vector $\vec{Y} = [Y_1 Y_2 \dots Y_n]^T$ that is linearly related to \vec{X} by

$$\vec{Y} = A\vec{X} + \vec{b}$$

where A is an $n \times n$ invertible matrix of constants and \vec{b} is an $n \times 1$ vector of constants.

- We will denote the inverse transformation by

$$\vec{X}(\vec{Y}) = A^{-1}\vec{Y} - A^{-1}\vec{b}$$

- The multivariate density of \vec{Y} is then given by

$$f_{\vec{Y}}(\vec{y}) = f_{\vec{X}}(\vec{X}(\vec{y})) \left| J\left(\frac{\vec{X}}{\vec{Y}}\right) \right|$$

where $\left| J\left(\frac{\vec{X}}{\vec{Y}}\right) \right|$ is the magnitude of the Jacobian

$$J\left(\frac{\vec{X}}{\vec{Y}}\right) = \det(A^{-1}) \quad \text{see book, page 54.}$$

- Thus, the density of \vec{Y} is given by

$$f_{\vec{Y}}(\vec{y}) = \frac{|\det A^{-1}|}{(2\pi)^{\frac{n}{2}} |C_{\vec{X}}|^{1/2}}$$

$$\times \exp \left\{ -\frac{1}{2} (\vec{A}^{-1}\vec{y} - \vec{A}^{-1}\vec{b} - \vec{m}_{\vec{X}})^T C_{\vec{X}}^{-1} (\vec{A}^{-1}\vec{y} - \vec{A}^{-1}\vec{b} - \vec{m}_{\vec{X}}) \right\}$$

Note: $E[\vec{Y}] = E[A\vec{X} + \vec{b}] = A\vec{m}_{\vec{X}} + \vec{b} = \vec{m}_{\vec{Y}}$

Note: $|\det A^{-1}| = \frac{1}{|\det A|} = \frac{1}{\sqrt{|\det A|} \sqrt{|\det A^T|}}$

- Then the density of \vec{Y} may be written as

$$f_{\vec{Y}}(\vec{y}) = (2\pi)^{\frac{n}{2}} |AC_{\vec{X}}A^T|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\vec{y} - \vec{m}_{\vec{Y}})^T (AC_{\vec{X}}A^T)^{-1} (\vec{y} - \vec{m}_{\vec{Y}}) \right\}$$

$\Rightarrow \vec{Y}$ is a multivariate normal vector with mean

$$\vec{m}_{\vec{Y}} = A\vec{m}_{\vec{X}} + \vec{b}$$

and covariance

$$C_{\vec{Y}} = AC_{\vec{X}}A^T.$$

\Rightarrow Given a jointly normal vector \vec{X} , it is of interest to find a linear transformation $\vec{Y} = A\vec{X} + \vec{b}$ such that the jointly Gaussian variables Y_1, Y_2, \dots, Y_n are mutually uncorrelated, and therefore mutually independent

\Rightarrow We say that such a transformation "decouples" the variables x_1, x_2, \dots, x_n .

\Rightarrow If y_1, y_2, \dots, y_n are mutually uncorrelated, then the covariance matrix C_y is diagonal:

$$C_y = \text{diag}(\sigma_{y_1}^2, \sigma_{y_2}^2, \dots, \sigma_{y_n}^2).$$

- Thus, a "decoupling" transformation (if it exists) diagonalizes the covariance matrix.

Note: a matrix A is positive definite if, $\forall \vec{b} \neq \vec{0}$,

$$\vec{b}^T A \vec{b} > 0,$$

and positive semidefinite if, $\forall \vec{b} \neq \vec{0}$,

$$\vec{b}^T A \vec{b} \geq 0.$$

FACT: Any covariance matrix is positive semidefinite.

FACT: If the magnitudes of all the correlation coefficients between the variables x_1, x_2, \dots, x_n are strictly less than unity, then

C_x is positive definite.

FACT: If C_x is positive definite, then a "decoupling" or "diagonalizing" transformation exists.

Facts About Jointly Gaussian Variables

Note: The list at the bottom of page 56 of the book needs two minor corrections to point (2.).

1. The density of a multivariate normal vector \vec{X} is completely specified by the mean vector and covariance matrix.
2. $C_{\vec{X}}$ is positive semidefinite. All of the correlation coefficients are less than or equal to one.
3. For jointly normal variables,
uncorrelated \longleftrightarrow independent
4. A linear transformation of a multivariate normal vector always yields another multivariate normal vector. A decoupling transformation exists if $C_{\vec{X}}$ is positive definite.
5. The marginal densities for X_1, X_2, \dots, X_n are all univariate normal densities.

Types of Convergence

- Suppose we have a sequence of RVS, Y_1, Y_2, \dots
- For example, let X_1, X_2, X_3, \dots be an infinite collection of iid RVS. We can define a sequence of RVS, Y_n by

$$Y_1 = X_1$$

$$Y_2 = \frac{1}{2}(X_1 + X_2)$$

⋮

$$Y_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n).$$

- The sequence Y_n is said to converge to θ in mean if

$$\lim_{n \rightarrow \infty} E[(Y_n - \theta)^2] = 0.$$

Note: θ could be deterministic or stochastic.

- The sequence Y_n is said to converge to θ in probability if, $\forall \epsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|Y_n - \theta| > \epsilon) = 0$$

\Rightarrow If Y_n converges in mean, then it converges in probability.

\Rightarrow The converse is not true.

Note: if X_1, X_2, \dots is an infinite collection of iid RV's with $E[X_k] = m_x$ and $\text{Var}(X_k) = \sigma_x^2$,

and if $Y_1 = X_1$,
 \vdots

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k,$$

then

$$\begin{aligned} E[Y_n] &= E\left[\frac{1}{n} \sum_{k=1}^n X_k\right] \\ &= \frac{1}{n} \sum_{k=1}^n E[X_k] \\ &= \frac{1}{n} \cdot n \cdot m_x = m_x. \end{aligned}$$

\Rightarrow In this case, we say that Y_n is an "unbiased estimator" of m_x .

Note: $\text{Var}(Y_n) = \frac{\sigma_x^2}{n}$ (Show this)